DOCUMENT RESUME

ED 213 732 TM 820 076

AUTHOR Wilcox, Rand R.
TITLE A Two-Stage Procedure for Selecting the Best of Several Binomial Populations; [and] Some Exact Sample Sizes for Comparing the Squared Multiple Correlation Coefficient to a Standard; [and] An Improved Decision-Theoretic Coefficient for Tests. Studies in Measurement and Methodology, Work Unit 3: Technical Adequacy of Tests.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Nov 79
GRANT OB-NIE-G-78-0213
NOTE 42p.; Some pages are marginally legible due to broken type.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Bayesian Statistics; *Correlation; Hypothesis Testing; *Mathematical Models; Probability; Research Problems; *Statistical Analysis; Testing Problems
IDENTIFIERS Binomial Distribution; *Sample Size

ABSTRACT

Three separate papers are included in this report. The first describes a two-stage procedure for choosing from among several instructional programs the one which maximizes the probability of passing the test. The second gives the exact sample sizes required to determine whether a squared multiple correlation coefficient is above or below a known constant, rather than testing the null hypothesis. The results reported here also give a lower bound to the probability of making a correct decision when observations are available. The third paper proposes a rescaling of the Bayes risk. The motivation for this new coefficient was to provide an index that has a large value when the Bayes risk is small, and that has a value in the closed interval between zero and one. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED213732

CSE
ARCHIVE COPY

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

DELIVERABLE - November 1979

STUDIES IN MEASUREMENT & METHODOLOGY

Work Unit 3:

TECHNICAL ADEQUACY OF TESTS

Rand R. Wilcox
Project Director

A TWO-STAGE PROCEDURE FOR SELECTING THE BEST OF SEVERAL BINOMIAL POPULATIONS

SOME EXACT SAMPLE SIZES FOR COMPARING THE SQUARED MULTIPLE CORRELATION COEFFICIENT TO A STANDARD

AN IMPROVED DECISION-THEORETIC COEFFICIENT FOR TESTS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G. S. Gray
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Grant #OB-NIE-G-78-0213

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California - Los Angeles

TM 820 074

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

A TWO STAGE PROCEDURE FOR SELECTING THE BEST OF SEVERAL BINOMIAL POPULATIONS

Rand R. Wilcox

Center for the Study of Evaluation

University of California, Los Angeles

ABSTRACT

In many situations in education and psychology it is desired to select from k binomial populations the one having the largest probability of success. For example, if a pass/fail type of test is administered at the end of an instructional program and if we have several methods of instruction from which to choose, it is desirable to choose the one which maximizes the probability of passing the test. This paper describes a two-stage procedure for accomplishing this goal. The solution is conservative in the sense that it is possible to determine the number of observations required so that the best population is chosen with probability at least P^*, $2^{-k} < P^* < 1$.

1. Introduction

Sobel and Huyett (1957) have examined the ranking and selection problem of determining the "best" of several binomial populations. More specifically, let $\pi_1,\ldots,\pi_k$ be k binomial populations having probability of success $p_1,\ldots p_k$, respectively. Let $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]}$ be the ordered values of the p_i's (i=1,...,k). The problem is to determine the population having the largest probability of success $p_{[k]}$. Let x_i be the observed number of successes from population π_i based on n observations. The decision rule is to choose the population yielding the largest x_i. In case of ties we choose at random a population from among those having the largest x_i. The goal is to choose n so that the probability of a correct selection (CS) is at least P^*, $2^{-k} < P^* < 1$. Symbolically we want

$$P(CS) \geq P^* \qquad (1.1)$$

Following Bechhofer (1954), Levy (1975a; 1975b) suggests an approximate solution to this problem which requires prior information about the value of $p_{[k]}$ as well as $p_{[k]}-p_{[k-1]}$. As noted by Wilcox (1978), Levy's approximate solution is readily modified so that information about $p_{[k]}$ and $p_{[k]}-p_{[k-1]}$ is not required.

We observe that several sequential procedures have been proposed for selecting the binomial population having probability of success $p_{[k]}$ (e.g., Bechhofer, Kiefer and Sobel, 1968; Sobel and Weiss, 1972; Raghunandanan, 1974). The advantage of these procedures is that they may require substantially fewer

observations so that (1.1) is satisfied. As pointed out by Cronbach and Gleser (1965, p. 70), however, there might be administrative difficulties when using a sequential statistical procedure in which case a two-stage procedure might be preferred. In this paper we describe a two-stage procedure for selecting the best of several binomial populations which combines the approach adopted by Sobel and Huyett (1957) with a modification of the subset selection procedure of Gupta and Sobel (1960).

2. A Modified Subset Selection Procedure

As already indicated, we have k binomial populations and we want to select the one having the largest probability of success, namely, $p_{[k]}$. It is assumed that there is no a priori information regarding p_i $(i=1,\ldots,k)$. Following Sobel and Huyett (1957) let $\delta^* > 0$ represent the largest value of $p_{[k]}-p_{[k-1]}$ that the experimenter believes worth detecting. That is, if $p_{[k]}-p_{[k-1]} < \delta^*$ we are indifferent as to which of the two corresponding binomial populations is chosen. The open interval $(p_{[k]}, p_{[k]}-\delta^*)$ is known as the indifference zone. If $p_{[k]}-p_{[1]} < \delta^*$, then any of the binomial populations will do. If, however, $p_{[k]}-p_{[k-1]} \geq \delta^*$ we want, with probability at least P^*, to choose the population associated with $p_{[k]}$. As δ^* approaches zero the required value of n so that (1.1) is satisfied gets indefinitely large. Thus, for practical purposes, the experimenter will want to guard against choosing δ^* too small.

The first stage of our two-stage solution is designed to eliminate from consideration some of the binomial populations. In the second stage we simply apply the solution of Sobel and Huyett (1957) to the remaining populations.

More specifically, suppose N observations are randomly sampled from population Π_i and the resulting number of observed successes is x_i where x_i is independent of x_j, $i \neq j$. The value of N is chosen primarily for convenience (cf. Somerville, 1974). Let $x_{max} = \max_i x_i$. We eliminate from consideration those populations for which

$$x_i < x_{max} - c \qquad (2.1)$$

where c, a constant to be determined, is chosen so that the best population is retained for further study with probability at least $(P^*)^{1/2}$. The reason for taking the square root of P^* will become apparent.

Gupta and Sobel have already considered this "elimination" or subset selection problem for the important special case in which there is no indifference zone. However, this procedure is inefficient for the present problem since in the second stage the investigator must specify δ^*. In particular, we might be able to determine the best binomial population with substantially fewer observations than would be required if the Gupta-Sobel solution is used. Accordingly, we modify the Gupta-Sobel subset selection procedure to include an indifference zone (i.e., a $\delta^* > 0$). For the normal analogue of our modification see Ryan and Antle (1976).

Let $x_{(i)}$ be the observed number of successes sampled from the binomial population having probability of success $p_{[i]}$. To obtain the appropriate

values for c we first observe that a lower bound on the probability of including $x_{(k)}$ in the selected subset is given when $p_{[1]}=p_{[2]}=\ldots=p_{[k-1]}=p$, say, and $p_{[k]}=p+\delta^*$. For this configuration of the k probabilities, the probability of including the best binomial population in the selected subset is

$$\sum_{x=0}^{N} \binom{N}{x}(p+\delta^*)(1-p-\delta^*)\{\sum_{j=0}^{x+c}\binom{N}{j} p^j(1-p)^{N-j}\}^{k-1} \qquad (2.2)$$

as shown by Gupta and Sobel. The appropriate choice for c is the smallest nonnegative integer so that (2.2) is greater than or equal to $(P^*)^{1/2}$ for any p.

We indicate an approximate solution for determining c. Gupta and Sobel approximate (2.2) with

$$\int_{-\infty}^{\infty} \{F\left(x\sqrt{\frac{(p+\delta^*)(1-p-\delta^*)}{p(1-p)}} + (\frac{c}{N}+\delta^*)\sqrt{\frac{N}{p(1-p)}}\right)\}^{k-1} f(x)dx \qquad (2.3)$$

where F(x) and f(x) are the cumulative distribution and probability density function of a standard normal random variable. They also point out that the minimum of (2.2) occurs at or near p=1/2. For p=1/2 and δ^* small, the value of (2.2) is given approximately by

$$\int_{-\infty}^{\infty} \{F(x+2(\frac{c}{N}+\delta^*)\sqrt{N})\}^{k-1} f(x)dx \qquad (2.4)$$

Thus, the appropriate value of c is the smallest nonnegative integer for which (2.4) is greater than or equal to $(P^*)^{1/2}$. Values of b for which

$$\int_{-\infty}^{\infty} \{F(x+b)\}^{k-1} f(x)dx=(P^*)^{1/2} \qquad (2.5)$$

are given by Bechhofer (1954) for k=2(1)10 and by Gupta (1956) as well as Gibbons, Olkin and Sobel (1977) for k=2(1)50. Once b is determined, it is sufficient to set

$$c = N(b(4N)^{-1/2} - \delta^*) \tag{2.6}$$

since for larger values of b the left hand side of (2.5) is greater than $(P^*)^{1/2}$ which implies that (2.4) is also.

As a partial check on the accuracy of (2.6) we determined c using (2.2) for $(P^*)^{1/2}$=.75, k=10, N=1(1)20(5)50(10)100. In every case we found, after rounding, exact agreement between c determined by (2.5) and c determined by (2.2). However, for $(P^*)^{1/2}$ close to one this approximation might not be completely satisfactory and so we calculated the values of c for δ^*=.05, .1, N=5(1)20(5)50, k=2(1)20,25, $(P^*)^{1/2}$=.95, .99. The results are reported in Table 1.

2.1 The Second Stage

As already indicated, the second stage consists of applying the solution described by Sobel and Huyett. We merely draw a new random sample of size n from each of the m remaining populations. Letting y_i represent the observed number of successes corresponding to the m populations, where y_i is independent of x_i, we choose the population yielding the maximum value of y_i. From the results reported by Sobel and Huyett it is possible to choose n so that in the second stage the probability of selecting the best population is at least $(P^*)^{1/2}$. Since by assumption x_i is independent of y_i, we choose the best binomial population with probability at least $(P^*)^{1/2}\,(P^*)^{1/2}=P^*$.

Another method of determining the best population is to choose the one yielding the largest value of $z_i=y_i+x_i$. Indeed, this is a reasonable manner in which to proceed. Our goal here, however, is to determine a lower bound on the probability of selecting the best of the k binomial populations based on the number of observations used. From Somerville (1974) we have that in the asymptotic case our procedure will give a lower bound to the P(CS) when z_i rather than y_i is used in the second stage. Somerville's results also provide some guidelines on how we might choose N and n given δ^* and P^*.

2.2 An Illustration

To illustrate the above procedure as well as its potential advantages, suppose we are considering k=10 binomial populations, $P^*=.9$, $\delta^*=.1$ and that unknown to us $p_{[1]}=p_{[2]}=\ldots=p_{[k-1]}=.65$ and $p_{[k]}=.90$. Under the single stage procedure we require ten sets of 222 observations. Suppose instead the investigator decides to use the two-stage procedure just described with N=25 in the first stage. In this case we set $(P^*)^{1/2}=.95$ (since $(.95)^2=.90$). From Bechhofer's Table 1 (the column headed k=10, t=1) b=3.4182 and so c=6.05 which after rounding becomes c=6. We see that this agrees with the exact value of c reported in Table 1. If we use the Gupta-Sobel procedure with no indifference zone ($\delta^*=0$), c=8 is required. Suppose that the observed number of successes for the ten sets of 25 observation is 15, 15, 15, 18, 21, 19, 17, 17, 15, 24. These numbers were generated with a random number generator and the values of $p_{[1]},\ldots,p_{[k]}$ given above. We retain for further

consideration those populations for which the observed number of successes is greater than or equal to $x_{max}-c=24-6=13$. Thus, we eliminate all but four of the populations. In the second stage we want to choose the best binomial population from those that remain with probability at least .95. From the tables given by Sobel and Huyett, we need four sets of 212 observations in the second stage. In this particular instance we required a total of 4(212) + 10(25) = 1098 observations which is approximately one-half the number required in the single-stage solution.

TABLE 1a

Values of c, $P^* = .95$, $\delta^* = .05$

k \ N	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	35	40	45	50
2	2	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	6	6
3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	6	6	6	7	7	7
4	3	3	3	4	4	4	4	4	5	5	5	5	5	5	5	5	6	6	7	7	8	8
5	3	3	4	4	4	4	4	5	5	5	5	5	5	6	6	6	6	7	7	8	8	8
6	3	3	4	4	4	4	5	5	5	5	5	5	6	6	6	6	7	7	8	8	8	9
7	3	4	4	4	4	5	5	5	5	5	5	6	6	6	6	6	7	7	8	8	9	9
8	3	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6	7	8	8	8	9	9
9	3	4	4	4	5	5	5	5	5	6	6	6	6	6	6	6	7	8	8	9	9	9
10	3	4	4	4	5	5	5	5	5	6	6	6	6	6	6	7	7	8	8	9	9	10
11	4	4	4	4	5	5	5	5	6	6	6	6	6	6	7	7	7	8	8	9	9	10
12	4	4	4	4	5	5	5	5	6	6	6	6	6	6	7	7	7	8	9	9	9	10
13	4	4	4	5	5	5	5	5	6	6	6	6	6	7	7	7	8	8	9	9	10	10
14	4	4	4	5	5	5	5	6	6	6	6	6	6	7	7	7	8	8	9	9	10	10
15	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	8	8	9	9	10	10
16	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	8	8	9	9	10	10
17	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	8	8	9	10	10	10
18	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	8	9	9	10	10	10
19	4	4	4	5	5	5	5	6	6	6	6	7	7	7	7	7	8	9	9	10	10	11
20	4	4	4	5	5	5	6	6	6	6	6	7	7	7	7	7	8	9	9	10	10	11
25	4	4	5	5	5	5	6	6	6	6	7	7	7	7	7	7	8	9	9	10	11	11

TABLE 1b

Values of c, $P^* = .99$, $\delta^* = .05$

k \ N	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	35	40	45	50
2	3	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6	7	7	8	8	9	9
3	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	8	8	9	9	10	10
4	4	4	5	5	5	5	6	6	6	6	7	7	7	7	7	7	8	9	9	10	10	11
5	4	4	5	5	5	6	6	6	6	7	7	7	7	7	8	8	8	9	10	10	11	11
6	4	4	5	5	5	6	6	6	6	7	7	7	7	8	8	8	9	9	10	11	11	12
7	4	5	5	5	6	6	6	6	7	7	7	7	7	8	8	8	9	10	10	11	11	12
8	4	5	5	5	6	6	6	6	7	7	7	7	8	8	8	8	9	10	10	11	12	12
9	4	5	5	5	6	6	6	7	7	7	7	7	8	8	8	8	9	10	11	11	12	12
10	4	5	5	5	6	6	6	7	7	7	7	8	8	8	8	8	9	10	11	11	12	12
11	4	5	5	5	6	6	6	7	7	7	7	8	8	8	8	8	9	10	11	12	12	13
12	4	5	5	6	6	6	6	7	7	7	7	8	8	8	8	9	9	10	11	12	12	13
13	4	5	5	6	6	6	7	7	7	7	8	8	8	8	8	9	10	10	11	12	12	13
14	4	5	5	6	6	6	7	7	7	7	8	8	8	8	9	9	10	10	11	12	12	13
15	4	5	5	6	6	6	7	7	7	7	8	8	8	8	9	9	10	11	11	12	13	13
16	4	5	5	6	6	6	7	7	7	7	8	8	8	8	9	9	10	11	11	12	13	13
17	4	5	5	6	6	6	7	7	7	8	8	8	8	8	9	9	10	11	11	12	13	13
18	4	5	5	6	6	6	7	7	7	8	8	8	8	8	9	9	10	11	11	12	13	13
19	4	5	5	6	6	6	7	7	7	8	8	8	8	8	9	9	10	11	11	12	13	13
20	4	5	5	6	6	6	7	7	7	8	8	8	8	8	9	9	10	11	11	12	13	13
25	5	5	5	6	6	7	7	7	7	8	8	8	9	9	9	9	10	11	12	13	13	14

TABLE 1c

Values of c, $P^* = .95$, $\delta^* = .10$

k \ N	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	35	40	45	50
2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	2	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	5	5	5
4	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5
5	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5	5	6	6	6	6
6	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6	6
7	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5	5	6	6	6	6	6	6
8	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5	5	6	6	6	6	7	7
9	3	3	4	4	4	4	4	5	5	5	5	5	5	5	5	5	6	6	6	7	7	7
10	3	4	4	4	4	4	5	5	5	5	5	5	5	5	6	6	6	6	7	7	7	7
11	3	4	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6	6	7	7	7	7
12	3	4	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6	7	7	7	7	7
13	3	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6	6	7	7	7	7	7
14	3	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6	6	7	7	7	7	8
15	3	4	4	4	4	5	5	5	5	5	5	6	6	6	6	6	6	7	7	7	8	8
16	3	4	4	4	4	5	5	5	5	5	5	6	6	6	6	6	7	7	7	7	8	8
17	3	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6	7	7	7	8	8	8
18	3	4	4	4	5	5	5	5	5	5	6	6	6	6	6	6	7	7	7	8	8	8
19	3	4	4	4	5	5	5	5	5	5	6	6	6	6	6	6	7	7	7	8	8	8
20	4	4	4	4	5	5	5	5	5	5	6	6	6	6	6	6	7	7	7	8	8	8
25	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6	6	7	7	8	8	8	8

TABLE 1d

Values of c, $P^* = .99$, $\delta^* = .10$

k \ N	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	35	40	45	50
2	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5	5	6	6	6	6	7	7
3	3	4	4	4	4	5	5	5	5	5	5	6	6	6	6	6	7	7	7	7	8	8
4	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6	6	7	7	8	8	8	8
5	4	4	4	5	5	5	5	6	6	6	6	6	6	6	7	7	7	8	8	8	9	9
6	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	7	8	8	9	9	9
7	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	7	8	8	9	9	9	9
8	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	7	8	8	9	9	9	10
9	4	4	5	5	5	6	6	6	6	6	7	7	7	7	7	7	8	8	9	9	10	10
10	4	4	5	5	5	6	6	6	6	6	7	7	7	7	7	7	8	9	9	9	10	10
11	4	4	5	5	5	6	6	6	6	6	7	7	7	7	7	8	8	9	9	10	10	10
12	4	4	5	5	5	6	6	6	6	6	7	7	7	7	7	8	8	9	9	10	10	10
13	4	5	5	5	5	6	6	6	6	7	7	7	7	7	8	8	8	9	9	10	10	10
14	4	5	5	5	6	6	6	6	6	7	7	7	7	7	8	8	8	9	9	10	10	10
15	4	5	5	5	6	6	6	6	7	7	7	7	7	7	8	8	8	9	9	10	10	11
16	4	5	5	5	6	6	6	6	7	7	7	7	7	8	8	8	9	9	10	10	10	11
17	4	5	5	5	6	6	6	6	7	7	7	7	7	8	8	8	9	9	10	10	10	11
18	4	5	5	5	6	6	6	6	7	7	7	7	7	8	8	8	9	9	10	10	11	11
19	4	5	5	5	6	6	6	6	7	7	7	7	7	8	8	8	9	9	10	10	11	11
20	4	5	5	5	6	6	6	6	7	7	7	7	8	8	8	8	9	9	10	10	11	11
25	4	5	5	6	6	6	6	7	7	7	7	8	8	8	8	8	9	10	10	11	11	11

Bechhofer, R. E. A single-sample multiple decision procedure for ranking means of normal populations with known variances. Annals of Mathematical Statistics, 1954, 25, 16-39.

Bechhofer, R. E., Kiefer, J., and Sobel, M. Sequential identification and ranking procedures. Chicago: University of Chicago Press, 1968.

Cronbach, L. J. and Gleser, G. C. Psychological tests and personnel decisions. (2nd Ed.) Urbana, Illinois: University of Illinois Press, 1965.

Gibbons, J., Olkin, I., and Sobel, M. Selecting and ordering populations: A new statistical methodology. New York: John Wiley, 1977.

Gupta, S. S. On a decision rule for a problem in ranking means. Institute of Statistics Mimeo Series No. 150, Institute of Statistics, University of North Carolina, Chapell Hill, 1956.

Gupta S. S. and Sobel, M. Selecting a subset containing the best of several binomial populations. In I. Olkin (Ed.) Contributions to probability and statistics, Stanford: Stanford University Press, 1960.

Levy, K. J. Appropriate sample sizes for selecting the population with the largest value of π from among k binomial populations. British Journal of Mathematical and Statistical Psychology, 1975, 28, 134-137(a).

Levy, K. J. Selecting the best population from among k binomial populations or the population with the largest correlation coefficient from among k bivariate normal populations. Psychometrika, 1975, 40, 121-122(b).

Raghunandanan, K. On play the winner sampling rule. Communications in Statistics, 1974, 3, 769-776.

Ryan, T. A. and Antle, C. E. A note on Gupta's selection procedure. Journal of the American Statistical Association, 1976, 71, 140-142.

Sobel, M. and Huyett, M. Selecting the best one of several binomial populations. Bell System Technical Journal, 1957, 36, 537-576.

Sobel and Weiss, G. H. Play-the-winner and inverse sampling for selecting the best of $k \geq 3$ binomial populations. Annals of Mathematical Statistics, 1972, 43, 1808-1826.

Somerville, P. N. On allocation of resources in a two-stage selection procedure. Sankhya, 1974, 36, Ser. B, 194-203.

Wilcox, R. R. Some comments on selecting the best of several binomial populations or the bivariate normal population having the largest correlation coefficient. Psychometrika, 1978, 43, 127-128.

Some Exact Sample Sizes for Comparing the Squared Multiple Correlation Coefficient to a Standard

Rand R. Wilcox

Center for the Study of Evaluation

University of California at Los Angeles

CSE
ARCHIVE COPY

The work upon which this publication is based was performed pursuant to a grant (contract) with the National Institue of Education, Department of Health, Education and Welfare. Points of view or opinions stated do not necessarily represent official NIE position or policy.

ABSTRACT

When analyzing ρ, the squared multiple correlation coefficient, an investigator may be more interested in determining whether ρ is above or below a known constant rather than testing the hypothesis that $\rho=0$. This paper gives the exact sample sizes required for answering this question when an indifference zone formulation of the problem is used. The results reported here also give a lower bound to the probability of making a correct decision when n observations are available.

In education and psychology, as well as a number of other disciplines, a common method of characterizing a set of observations is to use the squared multiple correlation coefficient. Typically the analysis of this coefficient consists of testing the hypothesis that its value is equal to zero. In some instances, however, an investigator may be more interested in determining whether the squared multiple correlation coefficient is "small" or "large" rather than determining that it is not equal to zero. This paper gives the exact sample sizes required for answering this question when an indifference zone formulation of the problem is used.

Formal Statement of the Problem

Consider a p-variate normal distribution and let ρ represent the squared multiple correlation coefficient between the first variate and the set of p-1 variates remaining. Let $\hat{\rho}$ be the usual estimate of ρ based on a random sample of n observations. Using the indifference zone approach described by Gibbons, Olkin and Sobel (1977, Chapter 10) this paper considers the problem of determining whether ρ is above or below the known constant ρ_0. The decision $\rho \geq \rho_0$ is made if $\hat{\rho} \geq \rho_0$; if $\hat{\rho} < \rho_0$, the reverse is said to be true. The goal is to choose the smallest n so that the probability of a correct decision is at least P^*, $2^{-k} < P^* < 1$, regardless of the true value of ρ.

When using the indifference zone approach to this problem, the investigator specifies the additional constant δ^*. The open interval $(\rho_0-\delta^*, \rho_0+\delta^*)$

is known as the indifference zone. If $\rho < \rho_0 - \delta^*$ or if $\rho > \rho_0 + \delta^*$ we want to be reasonably certain that a correct decision is made. If $\rho_0 - \delta^* < \rho < \rho_0 + \delta^*$, a correct decision is said to be made with probability one since an incorrect decision entails negligible loss.

Let $F_\rho(\hat{\rho})$ be the cumulative distribution function of the statistic $\hat{\rho}$. Recently, Alam, Rizvi and Solomon (1976) showed that the distribution of $\hat{\rho}$ is stochastically increasing in ρ. This means that $\rho < \rho'$ implies $F_{\rho'}(\hat{\rho}) \leq F_\rho(\hat{\rho})$ for all values of $\hat{\rho}$. It follows that for a given sample size n, the minimum probability of a correct decision is given by

$$F_{\rho=\rho_0-\delta^*}(\hat{\rho}) \qquad (1)$$

or

$$1 - F_{\rho=\rho_0+\delta^*}(\hat{\rho}) \qquad (2)$$

whichever is smallest. (See, e.g., Barr and Rizvi, 1966.) Alam, Rizvi, and Solomon also show that $\hat{\rho}$ has the monotone likelihood ratio property which also yields this result. Thus, to solve the present problem, it is sufficient to evaluate (1) and (2) for increasing values of n until both are greater than or equal to P^*. If $\delta^*=0$ (i.e., no indifference zone) it may be impossible to find a sufficiently large n.

Let $U=\hat{\rho}/(1-\hat{\rho})$. As shown by Gurland (1968), the cumulative distribution of U may be written as

$$F_U(x) = \sum_{j=0}^{\infty} \pi_j \, F_{p-1+2j,2k}(x)$$

where

$$\pi_j = \frac{(1-\hat{\rho})^{n/2}}{\Gamma(n/2)} \frac{\Gamma(n/2+j)}{\Gamma(j+1)} \hat{\rho}^j$$

and $F_{r,s}$ is the cumulative distribution function of a ratio χ_r^2 / χ_s^2 of independent chi-squared variables with degrees of freedom as indicated. F_U was used to evaluate (1) and (2) for increasing values of n until both exceeded P^*. The required sample size is reported in Table 1 for $\delta^*=.1$, $\rho_0=.15$, (.05) .85, p=2(1)10 and also $\delta^*=.05$, $\rho_0=.1$, .9 and p=2(1)10.

As an illustration, consider the following situation encountered by a colleague. It was desired to study the relationship between reading ability and four independent variables, namely, SES, quality of staff development (i.e., the quality of methods designed to improve teacher's skills), usefulness of staff development, and number of training sessions given and attended. Appropriate measures for all five variables were developed but the exact nature of these measures need not concern us here. As a measure of association between the first variable and the remaining four, the squared multiple correlation coefficient was estimated based on results with 40 classrooms, classrooms being used as the unit of analysis. The estimate was $\hat{\rho}=.43$. The investigator was interested in determing whether ρ was above or below .15. From Table 1 it can be seen that with n=61, $\delta^*=.1$ the PCD (probability of a correct decision) is at least .75. Thus, with n=40, the minimum PCD is less than .75. Note, however, that the actual PCD depends on the true but unknown value of ρ and so the the PCD may be higher than .75 even with n=40. The point is, as previously noted, that the sample sizes reported in Table 1 are

conservative in the sense that the PCD $\geq P^*$ regardless of the value of ρ. Also note that for $\rho_0=.1$, $\delta^*=.05$ and n=29, the minimum PCD is .75. Thus, if the investigator had wanted to compare ρ to .1 with n=40, the PCD is greater than .75.

TABLE 1

Exact Sample Sizes, $\delta^* = .1$

p^*	p_0	p: 2	3	4	5	6	7	8	9	10
.75		11	15	18	21	25	28	31	34	37
.90		25	30	34	37	41	45	48	52	55
.95	.85	38	42	46	50	54	58	62	65	69
.975		54	58	62	66	70	74	78	82	86
.99		67	71	75	80	84	88	92	96	100
.75		14	19	23	27	31	35	39	43	46
.90		36	41	46	50	55	59	64	68	72
.95	.80	54	59	64	69	74	79	84	88	93
.975		78	83	88	93	98	103	108	112	117
.99		97	102	107	113	118	123	128	133	138
.75		17	23	28	33	38	42	47	52	56
.90		46	52	58	64	69	75	80	85	90
.95	.75	72	78	84	90	95	101	106	112	117
.975		102	109	115	121	127	133	139	145	150
.99		128	135	141	147	154	160	165	171	177
.75		21	27	33	39	44	50	55	60	65
.90		57	64	71	78	84	90	96	102	108
.95	.70	89	96	103	110	117	123	130	136	142
.975		128	136	143	150	157	164	170	177	184
.99		161	168	175	184	190	198	204	211	217
.75		24	31	38	45	51	57	63	69	75
.90		68	76	84	91	99	106	113	120	126
.95	.65	106	115	123	130	138	145	153	160	167
.975		154	162	171	179	186	194	201	209	217
.99		193	203	210	218	227	235	242	250	257
.75		27	35	43	50	57	64	71	77	84
.90		78	87	96	104	112	120	128	136	143
.95	.60	123	132	141	150	158	166	174	183	190
.975		178	186	196	205	214	222	231	240	248
.99		223	233	243	252	260	269	278	287	295
.75		29	38	47	55	63	71	78	85	93
.90		87	97	106	116	125	134	142	151	159
.95	.55	137	147	157	166	176	185	194	203	212
.975		198	209	219	229	238	248	257	266	276
.99		250	260	270	280	291	301	311	319	330
.75		31	41	51	59	68	76	85	93	101
.90		94	105	115	125	135	145	154	164	173
.95	.50	148	159	170	181	191	201	211	221	230
.975		215	226	237	248	260	270	280	291	301
.99		272	283	294	306	216	326	337	347	357

(i) Continued →

p^*	p_0	p: 2	3	4	5	6	7	8	9	10
.75		32	43	54	63	72	82	90	99	108
.90		98	110	122	133	144	154	164	175	184
.95	.45	156	168	180	191	203	214	225	235	246
.975		227	239	252	264	275	286	298	309	319
.99		287	298	312	322	334	346	356	369	379
.75		33	45	56	66	76	86	95	105	114
.90		101	114	126	138	149	161	172	183	193
.95	.40	160	173	185	198	210	222	233	245	256
.975		233	246	260	272	284	296	309	320	333
.99		294	306	319	333	345	356	369	382	393
.75		38	45	57	68	78	89	99	109	119
.90		100	114	127	140	152	164	176	187	198
.95	.35	158	172	186	199	212	225	237	250	261
.975		230	245	259	272	286	299	312	324	337
.99		291	305	319	333	347	360	374	385	398
.75		44	51	60	70	80	91	101	112	122
.90		95	110	124	137	150	163	175	187	199
.95	.30	150	165	180	194	207	221	234	247	260
.975		220	235	250	264	278	292	306	319	333
.99		276	292	307	321	336	351	365	377	393
.75		28	42	55	68	79	91	102	113	124
.90		86	102	117	131	145	158	171	184	196
.95	.25	137	153	168	183	198	212	225	239	252
.975		199	216	231	246	261	276	291	304	318
.99		250	266	283	298	313	329	342	357	372
.75		25	39	53	65	77	89	101	112	124
.90		72	89	105	119	134	148	161	174	187
.95	.20	115	132	148	164	179	193	207	221	235
.975		167	184	201	217	233	248	263	277	292
.99		210	228	245	260	276	292	307	322	338
.75		21	34	48	61	73	86	97	109	121
.90		80	82	87	102	117	131	144	158	171
.95	.15	83	102	119	135	150	165	180	194	208
.975		121	140	157	174	190	205	221	235	251
.99		153	171	189	205	222	238	254	269	284

Continued ⟶

Exact Sample Sizes, $\delta^* = .05$

p^*	p_0	p: 2	3	4	5	6	7	8	9	10
.75		65	83	112	139	166	191	216	241	265
.90		164	200	234	266	296	326	355	384	412
.95	.10	264	301	335	369	401	433	464	494	524
.975		387	424	460	495	529	561	593	625	656
.99		490	524	562	598	600	665	697	730	764
.75		16	21	25	29	33	37	41	45	49
.90		41	46	51	56	61	65	70	74	78
.95	.90	63	68	73	78	83	88	93	97	102
.975		91	96	101	106	112	117	121	126	131
.99		115	120	126	131	136	141	146	151	156

REFERENCES

Alam, K., Rizvi, M. H. and Solomon, H. Selection of largest multiple correlation coefficients: exact sample size case. The Annals of Statistics, 1976, 4, 614-620.

Barr, D. R. and Rizvi, M. H. An introduction to ranking and selection. Journal of the American Statistical Association, 1966, 61, 640-646.

Gibbons, J., Olkin, I., and Sobel, M. Selecting and ordering populations: A new statistical methodology. New York: John Wiley, 1977.

Gurland, J. A relatively simple form of the distribution of the multiple correlation coefficient. Journal of the Royal Statistical Society, 1968, Series B, 30, 276-283.

An Improved Decision-Theoretic Coefficient for Tests

Rand R. Wilcox

Center for the Study of Evaluation

University of California at Los Angeles

The work upon which this publication is based was performed pursuant to a grant (contract) with the National Institute of Education, Department of Health, Education and Welfare. Points of view or opinions stated do not necessarily represent official NIE position or policy.

ABSTRACT

Recently it was suggested that the Bayes risk might be used to characterize tests. To conform to common practices about indices, a rescaling of the Bayes risk was proposed. The motivation for this new coefficient, δ, was to provide an index that has a large value when the Bayes risk is small and that has a value in the closed interval [0, 1]. However, δ has neither of these properties. A modification of δ, called γ, is described that corrects both of these difficulties.

Recently, van der Linden and Mellenbergh (1978) urgued in favor of using decision-theoretic techniques for characterizing a test. As they point out, the Bayes risk is a natural index to use from this point of view. Van der Linden and Mellenbergh also note that the Bayes risk has two disadvantages. First, it is conventional in test theory to define indices so that the scale has a direction opposite to that in which the Bayes risk is represented. Second, although in test theory indices are nearly always defined on the closed interval [0, 1], the range of the possible values of the Bayes risk can be different.

Let $L(\hat{\theta}(x)|\theta)$ be the loss function associated with a particular decision problem which is a function of the decision rule $\hat{\theta}(x)$, the examinee's true score, θ, and the examinee's observed score, x. For a particular examinee the risk or expected loss is $R = EL(\hat{\theta}(x)|\theta)$. The Bayes risk is $R_B=E_\theta R$, where E_θ means expectation over the population of examinees. In an attempt to correct the two deficiencies described above, van der Linden and Mellenbergh suggest the index

$$\delta = (R_n-R_B)/(R_n-R_c)$$

where R_c and R_n are the Bayes risks in the situation in which the test contains, respectively, complete and no information about the true score. They point out, however, that the index δ does not correct the second difficulty; δ might not be in the interval [0, 1] since it can have a negative value. The purpose of this note is to indicate a simple modification of δ that solevs this problem. As indicated below, δ can also be equal to zero when $R_B=0$ which is the exact opposite of what was intended. This problem is corrected as well.

Coefficient γ

To ensure that δ has a value between 0 and 1 we replace R_n with R_u, the least upper bound on R_B and we replace R_c with R_t, the greatest lower bound on R_B. In this case δ becomes

$$\gamma = (R_u - R_B)/(R_u - R_t).$$

By requiring R_u to be the least upper bound on R_B, rather than just any upper bound, we guarantee that it is possible for γ to equal 0. By also choosing R_t in the manner described above, we guarantee the possibility of $\gamma=1$. It is evident that $0 \leq \gamma \leq 1$ as was required. For many practical situations R_u and R_t can be evaluated. We illustrate this point by reexamining the situations considered by van der Linden and Mellenbergh.

Point Estimation

Suppose the purpose of a test is to estimate true score, that $\hat{\theta}(x)$ is a decision rule for estimating θ based on the observed score x and that squared error loss is used, i.e., $L=(\hat{\theta}(x)-\theta)^2$. If, as is usually the case, the observed scores can have the values $0,1,\ldots,n$, n being the number of items on the test, the Bayes risk is given by

$$R_B = \sum_{x=0}^{n} \int_0^1 (\hat{\theta}(x)-\theta)^2 f(x|\theta)g(\theta)d\theta$$

where $f(x|\theta)$ is the conditional probability function of observed scores and $g(\theta)$ is the density function of true scores.

The least upper bound of R_B depends on such things as the assumptions made in a particular testing model, the instructional history of the examinees, the statistic $\hat{\theta}(x)$ and the number of items on the test. We first

consider a situation in which a minimum number of assumptions are made. We then examine the case where a two-term approximation to the compound-binomial error model is used. This latter model is sometimes adopted when an item sampling model is thought to be appropriate (Lord and Novick, 1968, Chapter 23).

For convenience we assume $0 \leq \theta \leq 1$. This is the case for the compound binomial error model. For situations where θ is not in this range, a rescaling is readily made so that this will be true. This rescaling, will in general, have no effect on the value of γ. The only motivation for requiring $0 \leq \theta \leq 1$ is to simplify the notation. We also assume $0 \leq \hat{\theta}(x) \leq 1$.

Except for highly unusual circumstances, it will be the case that $\hat{\theta}(x)$ is a monotonically increasing function of x. It follows that the least upper bound of $(\hat{\theta}(x)-\theta)^2$ is either $(\hat{\theta}(0)-1)^2$ or $(\hat{\theta}(1)-0)^2$ whichever is largest. We let L_2 denote the larger of these two quantities. We have that $R \leq L_2$ since L_2 is an upper bound to $(\hat{\theta}(x)-\theta)^2$. For the case $L_2 = (\hat{\theta}(0)-1)^2$, $R=L_2$ when $Pr(x=0, \theta=1)=1$. If $L_2=(\hat{\theta}(1)-0)^2$, $R_B=L_2$ when $Pr(x=n, \theta=0)$. Note that this least upper bound was derived under the assumption that any joint probability density function for θ and x is possible. If we assume that $E(\hat{\theta}(x)|\theta)=\theta$ then in general L_2 will not be the least upper bound. Such a case is considered below.

The greatest lower bound for R_B is zero. This occurs, for example, when $Pr(x=b)=Pr(\theta=\hat{\theta}(b)=1$, where b is any integer, $0 \leq b \leq n$. Thus, γ may be written as

$$\gamma = (L_2-R)/L_2$$

Before considering more specific cases we indicate a serious difficulty with δ. Van der Linden and Mellenbergh use Kelley's regression equation to estimate true score, namely,

$$\theta(x) = \rho_{xx'}(x/n) + (1-\rho_{xx'})E(x/n) \qquad [1]$$

where $\rho_{xx'}$ is the reliability of the test. From results reported by Griffin and Krutchkoff (1971) it can be seen that this is the optimal linear estimator of true score has a smaller Bayes risk. In this case

$$R_B = \sigma_\theta^2(1-\rho_{xx'})$$

and so R_B can be estimated using known techniques. Van der Linden and Mellenbergh also show that when [1] is used to estimate true score, δ is equal to $\rho_{xx'}$. The inadequacy of δ as an index for tests can now be demonstrated. Suppose $\sigma_\theta^2=0$ in which case $R_B=0$. Thus, in terms of parameters, Kelley's regression equation gives the exact value of an examinee's true score. However, $\sigma_\theta^2=0$ implies that $\rho_{xx'}=0$ which means that $\delta=0$. Thus, van der Linden and Mellenbergh's coefficient δ indicates that the test used in conjunction with $\hat{\theta}(x)$ is a poor measuring instrument of true score when in fact the exact opposite is true. In contrast, $\gamma=1$ when $R_B=0$ as we would want.

As previously indicated, the derivation of L_2 was made without any restriction on the joint probability density function of x and θ or the decision rule $\hat{\theta}(x)$. In general, if we assume that $E(\hat{\theta}(x)|\theta)=\theta$, L_2 is no longer the least upper bound to R_B. This means that $w \geq d > 0$ for some number d. We indicate a solution to this problem for the important special case where the conditional distribution of observed scores is given by a two-term approximation

to the compound binomial distribution. From Lord (1965) we see that for $\hat{\theta}(x)=x/n$,

$$E(\hat{\theta}(x)|\theta)=\theta$$

and

$$E[(\hat{\theta}(x)-\theta)^2|\theta] = (n-2k)\theta(1-\theta) \qquad [2]$$

where $k \geq 0$ is given by

$$k = \frac{n^2(n-1)\sigma_\pi^2}{2[\mu_x(n-\mu_x)-\sigma_x^2-n\sigma_\pi^2]},$$

where σ_π^2 is the variance of the item difficulties, μ_x and σ_x^2 are the mean and variance of the observed scores. From [2] it is readily verified that

$$E[(\hat{\theta}(x)-\theta)^2|\theta] \leq 1/4n$$

with equality holding when $\theta=1/2$. Taking the expectation of both sides of [3], we see that

$$E_\theta[(\hat{\theta}(x)-\theta)^2] \leq 1/4n.$$

Thus, $1/4n$ is the least upper bound of R_B. Substituting $1/4n$ for R_u in γ we have that

$$\gamma = (1-4nR)/4n$$

since the greatest lower bound is again zero. Griffin and Krutchkoff indicate that for $\hat{\theta}(x)=x/n$, $R=\sigma_{x/n}^2-\sigma_\theta^2$ and so R can be estimated using results given by Lord and Novick (1968, Chapter 23).

If we replace the sample mean estimate of true score with Griffin and Krutchkoff's optimal linear estimator, then

$$R_B = \frac{\sigma_\theta^2}{\sigma_{x/n}^2}[\sigma_{x/n}^2 - \sigma_\pi^2].$$

It is not immediately clear what the least upper bound of R is in this case. However, a good approximation to it can be obtained by evaluation $E[(\hat{\theta}(x)-\theta)^2|\theta]$ for θ=.01, .02,...,.99 and setting R_u equal to the largest resulting value. This is easily done on a computer.

Dichotomous Decisions

Consider the problem of determining whether an examinee's true score θ is above or below a known constant θ_0. The decision $\theta \geq \theta_0$ is made if $x \geq x_0$ where x_0 is a specified passing score. This problem has been considered by several authors in the context of a mastery test (Harris, 1974; Novick and Lewis, 1974; Huynh, 1976; Wilcox, 1977). Let ℓ_{10} be the loss incurred when $x \geq x_0$ and $\theta < \theta_0$. Let ℓ_{01} be the loss when $x < x_0$ and $\theta \geq \theta$ If a correct decision is made we assume the loss is zero. For a randomly selected examinee, the probabilities p_{ij} (i=0, 1; j=0, 1) associated with the four possible outcomes are shown in Table 1.

TABLE 1

Probabilities Associated with the Four Possible Outcomes

True Score	$x < x_0$	$x \geq x_0$
$\theta < \theta_0$	P_{00}	P_{01}
$\theta \geq \theta_0$	P_{10}	P_{11}

Thus, p_{10} is the probability of $\theta \geq \theta_0$ and $x < x_0$. The Bayes risk is

$$R_B = \ell_{10} p_{10} + \ell_{01} p_{01}.$$

As before, the least upper bound of R is highly dependent upon the situation at hand. We consider several cases.

First note that it is always possible to guarantee that the probability of a correct decision for a given examinee is at least 1/2 simply by making the decision $\theta \geq \theta_0$ or $\theta < \theta_0$ at random. When this decision rule is used, the least upper bound of R_B is

$$R_u = \ell_m / 2$$

where $\ell_m = \max[\ell_{10}, \ell_{01}]$. Thus

$$\gamma = (\ell_m - 2R)/\ell_m$$

Next suppose that $f(x|\theta)$ is stochastically increasing. This means that $\theta < \theta'$ implies that $F(x|\theta) \leq F(x|\theta')$ for all x where $F(x|\theta)$ is the cumulative distribution corresponding to $f(x|\theta)$. The binomial and Poisson distributions are two examples where this is true. For such distributions the greatest lower bound of the probability of a correct decision for a given examinee is $\sum_{x=0}^{x_0-1} f(x|\theta_0)$ of $\sum_{x=x_0}^{n} f(x|\theta)$, whichever is smallest (Wilcox, in press). Denote this quantity by p_m. Then the least upper bound of R_B is $\ell_m(1-p_m)$ and so

$$\gamma = [\ell_m(1-p_m) - R]/[\ell_m(1-p_m)] \qquad [5]$$

Note that p_m may not equal 1/2 but usually, as n gets large, p_m approaches 1/2 which may be seen via the central limit theorem.

As a final example we follow Fhaner (1974) and Wilcox (in press) and consider the problem of deciding whether θ is below θ_1 or above θ_2, $\theta_1<\theta_2$. If $\theta_1<\theta<\theta_2$ either decision is said to be correct. The open interval (θ_1,θ_2) is called the indifference zone. As noted by Wilcox, the least upper bound on the probability of an incorrect decision is either $\sum_{x=0}^{x_0-1} f(x|\theta_2)$ or $\sum_{x=x_0}^{n} f(x|\theta_1)$, whichever is largest. Denoting this quantity as p_m, γ is given by [5].

Concluding Remarks

Coefficient γ was intended to correct two deficiencies that are associated with the Bayes risk as a description of a test. In actuality it corrects neither one since it can have the value zero when $R_B=0$ and since it can have values outside the closed interval [0, 1]. The first difficulty is easily corrected once an upper bound to R_B is found. If both a least upper bound and the greatest lower bound can be determined, the index γ can be used which always has a value between 0 and 1. Moreover, all values between 0 and 1 are possible.

REFERENCES

Fhaner, Stig. Item sampling and decision making in achievement testing. British Journal of Mathematical and Statistical Psychology, 1974, 27, 172-175.

Griffin, B. S. and Krutchkoff, R. G. Optimal Linear estimators: an empirical Bayes version with application to the binomial distribution. Biometrika, 1971, 58, 195-201.

Harris, C. W. Some technical characteristics of mastery tests. In C. W. Harris, M. C. Alkin and W. James Popham (Eds.) Problems in criterion-referenced measurement. CSE Monograph No. 3, Los Angeles: Center for the Study of Evaluation, University of California, 1974.

Huynh, H. Statistical consideration of mastery scores. Psychometrika, 1976, 41, 65-78.

Lord, F. M. A strong true-score theory, with applications. Psychometrika, 1965, 30, 239-270.

Lord, F. M. and Novick, M. R. Statistical theories of mental test scores. Reading, Mass: Addison - Wesley, 1968.

Novick, M. R. and Lewis, C. Prescribing test length for criterion-referenced measurement. In C. W. Harris, M. C. Alkin and W. James Popham (Eds.) Problems in criterion-referenced measurement. CSE Monograph No. 3, Los Angeles: Center for the Study of Evaluation, University of California, 1974.

Skibinsky, M. The maximum probability on an interval when the mean and variance are known. Sankhya, Series A, 1977, 39, 144-159.

van der Linden, W. J. and Mellenbergh, G. J. Coefficients for tests from a decision theoretic point of view. Applied Psychological Measurement, 1978, 2, 119-134.

Wilcox, R. R. Applying ranking and selection techniques to determine the length of a mastery test. Educational and Psychological Measurement, in press.

Wilcox, R. R. Estimating the likelihood of a false-positive on false-negative decision with a mastery test: an empirical Bayes approach. Journal of Educational Statistics, 1977, 2, 289-307.